A dissection of the cauliflower mosaic virus polyadenylation signal

Hélène Sanfaçon,[1] Peter Brodmann, and Thomas Hohn

Friedrich Miescher Institut, CH-4002, Basel, Switzerland

Mutagenesis analysis of the polyadenylation [poly(A)] signal from the cauliflower mosaic virus (CaMV), a plant pararetrovirus, revealed striking differences to known vertebrate poly(A) signals. Our results show that (1) the AATAAA sequence is necessary for efficient cleavage at the poly(A) site, although the requirement for an authentic AATAAA might be less stringent in plant than in vertebrate cells; (2) surprisingly and in contrast to the majority of vertebrate poly(A) signals, the sequences downstream of the CaMV poly(A) site do not influence processing efficiency drastically although they affect the precision of cleavage; and (3) deletion of sequences upstream of the CaMV AATAAA sequence decreased processing at the CaMV site dramatically, suggesting the presence of one or several positively acting upstream elements. An oligonucleotide consisting of CaMV upstream sequences could induce the recognition of a normally silent exogenous poly(A) signal when inserted upstream of its AATAAA motif.

[*Key Words*: mRNA 3'-end formation; plant polyadenylation signal; cauliflower mosaic virus; AATAAA mutation; upstream–downstream deletions; RNase protection assay]

Received July 30 1990; revised version accepted October 15, 1990.

The formation of the mRNA 3' end has been studied extensively in vertebrate cells (for recent reviews, see Humphreys and Proudfoot 1988; Manley 1988). The premature transcript is cleaved by an endonuclease prior to transcription termination, and a poly(A) tail is added at the cleavage site by the poly(A) polymerase. The cleavage and the poly(A) addition reactions are tightly coupled in vivo and involve large polyadenylation complexes containing many additional factors (Humphrey et al. 1987; Zarkower and Wickens 1987; Zhang and Cole 1987; Christofori and Keller 1988, 1989; Gilmartin et al. 1988; McDevitt et al. 1988; Takagaki et al. 1988). Efficient transcription termination depends on the presence of a functional poly(A) site (Whitelaw and Proudfoot 1986; Logan et al. 1987; Connelly and Manley 1988; Lanoix and Acheson 1988) and can occur up to several thousand nucleotides farther downstream (for review, see Proudfoot 1989) most likely at a polymerase II pause site (Connelly and Manley 1989).

In vertebrate genes, the hexanucleotide AATAAA is almost always present 10–30 nucleotides upstream of the cleavage site (Proudfoot and Brownlee 1976). Point mutations in this sequence greatly reduce the efficiency of 3'-end formation in vivo (Fitzgerald and Shenk 1981; Montell et al. 1983; Wickens and Stephenson 1984) and of both cleavage and polyadenylation in vitro (Manley et al. 1985; Zarkower et al. 1986; Conway and Wickens 1987; Wilusz et al. 1989). In addition, a less well-conserved T-rich or TG-rich element, usually situated immediately downstream of the poly(A) addition site, is important for both reactions in vivo and in vitro (Gil and Proudfoot 1984; Coles and Stacy 1985; Conway and Wickens 1985; Hart et al. 1985; McLauchlan et al. 1985; Sadofsky et al. 1985; Sperry and Berget 1986). Both the AATAAA and the downstream element are required for formation of the specific cleavage and polyadenylation complexes (Skolnik-David et al. 1987; Gilmartin and Nevins 1989). Additional upstream sequences were shown to improve recognition of the SV40 late poly(A) signal (Carswell and Alwine 1989) and the hepatitis B virus poly(A) signal (Russnak and Ganem 1990) and to influence poly(A) site selection in the adenovirus late transcription unit (DeZasso and Imperiale 1989).

Much less is known about mRNA 3'-end formation in plants. Polyadenylation signals from mammalian genes are not properly recognized in plants (Hunt et al. 1987), suggesting that either the sequences required for 3'-end formation or the mechanism of 3'-end formation might differ in plants. Heterogeneity of plant mRNA 3' ends due to the presence of multiple poly(A) signals has been described (Dean et al. 1986), contrasting with the single poly(A) site present in most vertebrate genes. Furthermore, a perfectly conserved AATAAA sequence is present in only 40% of known plant genes (Joshi 1987), suggesting that this sequence might not be absolutely required for 3'-end formation in plant cells. On a computer survey, a putative consensus sequence, TGTGTTT, was proposed to be involved in 3'-end formation. However, this sequence was found to be located in a widespread area downstream of the cleavage site (Joshi 1987). Very few functional analyses have been

[1]Present address: Agriculture Canada Research Station, Vancouver, B.C., V6T 1X2 Canada.

made using plant poly(A) signals. Analysis of the potato wound-inducible proteinase inhibitor II gene (An et al. 1989) and of the octopine synthetase gene (Ingelbrecht et al. 1989) showed reduction of reporter gene activity and of the steady-state RNA level on deletion of an AATAAA sequence and of a downstream TG-rich element. Deletion analysis of the pea ribulose-1,5-biphosphate carboxylase small-subunit gene suggested that elements located both upstream and downstream of the cleavage site are involved in mRNA 3′-end formation (Hunt and McDonald 1989).

The CaMV, a plant pararetrovirus, produces two major transcripts that are polyadenylated at the same site (Covey et al. 1981; Guilley et al. 1982). In a previous study, we showed that proximity to the promoter inhibits recognition of this signal. However, very efficient cleavage at the CaMV processing site was directed by the sequences contained in a 195-nucleotide-long CaMV fragment when placed downstream of a reporter gene (Sanfaçon and Hohn 1990). The primary structure of this fragment is shown in Figure 1. It spans 180 nucleotides upstream and 15 nucleotides downstream of the cleavage site. Here, we present a mutational analysis of this signal. Several elements involved in 3′-end formation were defined.

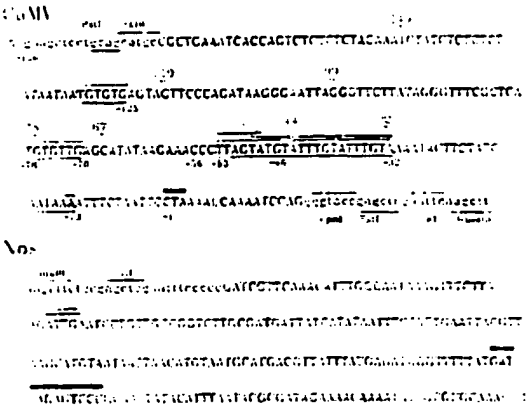


Figure 1. Primary structure of the CaMV and the *nos* polyadenylation signal fragments. Uppercase letters show original sequence from each signal; lowercase letters are polylinker sequences. The restriction sites are indicated by a thin line above or below the sequence. In the standard construct (R-CAT), the CaMV poly(A) signal fragment is followed immediately by the *nos* poly(A) signal fragment. The *Hin*dIII site is at the junction of those two fragments and is shown both at the end of the CaMV fragment and at the beginning of the *nos* fragment. Solid boxes above the sequence show the normal processing sites in the CaMV and *nos* fragments. The open box above the *nos* sequence represents a cryptic processing site (cryptic *nos* in the text). Sequences that may be important for polyadenylation are boxed: three upstream TG-rich sequences and the AATAAA sequence in the CaMV fragment. An AATAAA sequence that is not recognized in the wild-type *nos* gene (cryptic AATAAA in the text) is enclosed in a dotted box. Solid or open arrows above the third TG-rich box in the CaMV sequence show perfect and imperfect repeats of the sequence TTTGTA. The large numbers above the CaMV sequence show the exact end point of the upstream BAL-31 deletions discussed in the text; the smaller numbers below the sequence represent the position of relevant nucleotides relative to the CaMV processing site.

Results

Analysis of the primary structure of a 195-nucleotide CaMV fragment containing the poly(A) signal

A perfectly conserved AATAAA sequence is situated 13 nucleotides upstream of the poly(A) addition site of CaMV RNA (Fig. 1). Two TG-rich regions, resembling the TG-rich sequence found downstream of the poly(A) addition site in vertebrate genes (Gil and Proudfoot 1984; Coles and Stacy 1985; Conway and Wickens 1985; Sadofsky et al. 1985) and possibly in plant genes (Joshi 1987), are found 70 and 124 nucleotides upstream of the processing site. A third TG-rich region is located 32 nucleotides upstream of the processing site and consists of several tandem repeats of the sequence TTTGTA (Fig. 1). This sequence also resembles a sequence proposed to be involved in mRNA 3′-end formation in yeast (TAG . . . TAGT . . TTT; Zaret and Sherman 1982). Surprisingly, in contrast to vertebrate genes, the CaMV sequences downstream of the processing site are AC-rich.

An AATAAA sequence is required for processing at the CaMV site

The plasmid R-CAT (Sanfaçon and Hohn 1990; Fig. 2), containing the chloramphenicol acetyltransferase (CAT) reporter gene between CaMV 35S promoter and poly(A) signal, was used as a standard plasmid to study the CaMV poly(A) signal by site-directed mutagenesis. To rescue the RNA molecules that are not processed at the CaMV site, R-CAT contains a second poly(A) signal from the nopaline synthetase (*nos*) gene (Bevan et al. 1982; Depicker et al. 1982) downstream of the CaMV fragments. The proportion of transcripts processed at the CaMV versus the *nos* site was used as a measure for the efficiency of processing at the CaMV site. Plasmid R-CAT was introduced into *Nicotiana plumbaginifolia* protoplasts using a transient expression system, and total RNA was analyzed in an RNase protection assay with an antisense probe (P1) covering both poly(A) signals (see Fig. 2A). The standard plasmid gave rise to transcripts cleaved only at the CaMV poly(A) addition site, which protect 190 nucleotides of the probe (Fig. 3).

As a first step in the analysis of the CaMV poly(A) signal, the AATAAA sequence was mutated by oligonucleotide-directed mutagenesis of plasmid R-CAT. The precise deletion of the CaMV AATAAA prevented processing at the CaMV site (Fig. 3, construct ΔAATAAA), confirming our earlier results (Sanfaçon and Hohn 1990). Surprisingly, the rescued transcripts were not cleaved at the previously described *nos* processing site (Bevan et al. 1982; Depicker et al. 1982), but at a cryptic site ~100 nucleotides upstream and 20 nucleotides downstream of

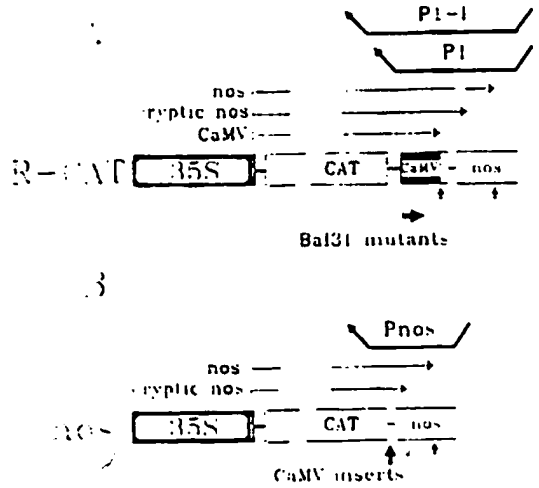


Figure 2. Strategy for the RNA mapping. (*A*) Strategy for the mapping of RNAs produced from the construct R-CAT or mutant derivatives. Boxes represent the CaMV 35S promoter enhancer, the CAT-coding region, a fragment containing the CaMV poly(A) signal, and a fragment containing the *nos* poly(A) signal. The thin lines above the construct represent the transcripts processed at either the CaMV site, the cryptic *nos*, or the normal *nos* sites (which are indicated by small arrows at *bottom*). The two antisense probes that were used with those constructs are indicated by the upper bold lines. The lower horizontal arrow indicates the starting point and the direction of the BAL-31 deletions. (*B*) Strategy for mapping of RNAs produced from the construct *nos* and its derivatives with CaMV inserts. Symbols are as in *A*. The site of insertion of the oligonucleotides containing specific CaMV sequences is indicated by the large solid arrow at the *bottom*.

an AATAAA sequence that is not recognized in the wild-type *nos* gene (Fig. 1; Bevan et al. 1982; Depicker et al. 1982).

A T → C point mutation of the AATAAA motif was recognized at half of the wild-type efficiency (Fig. 3, construct AAGAAA). Fifty-five percent of the transcripts were processed at the CaMV site and most of the remaining transcripts at the cryptic *nos* site (35%). It is noteworthy that in vertebrate in vivo (Montell et al. 1983) and in vitro (Wilusz et al. 1989) systems, a similar mutation would reduce cleavage by more than 95%.

These results suggest that the AATAAA sequence is required for efficient 3′-end processing in CaMV. Furthermore, additional sequences in the CaMV poly(A) signal are likely to be needed. These sequences probably allowed recognition of the cryptic AATAAA in the *nos* signal in absence of the CaMV AATAAA motif.

Sequences immediately downstream of the processing site influence the position of the CaMV cleavage site

A stretch of only 15 nucleotides of original CaMV sequence downstream of the processing site are included in plasmid R-CAT (see Fig. 1). Addition of the following 200 nucleotides farther downstream, isolated from the CaMV leader, had no effect on the efficiency of processing at the CaMV site (not shown). Therefore, if an important element downstream of the CaMV processing site exists, as in vertebrate genes, it must be located within those 15 nucleotides. To test this, this sequence was deleted precisely (Fig. 4, construct ΔDo) with the exception of an A residue close to the processing site, which would be of importance in vertebrate genes (Fitzgerald and Shenk 1981; Mason et al. 1986). Approximately 60% of the RNA molecules produced from this construct were cleaved at the CaMV site. The remainder were processed at either the cryptic *nos* site or at a new site upstream of it (Fig. 5). The slight decrease of processing at the CaMV site could be explained by two possibilities. We might have either deleted a positively acting element (e.g., an AC-rich region) or created a negatively acting element (e.g., by bringing the polylinker closer to the processing site).

To distinguish between these two possibilities, we further deleted most of the polylinker in either the standard plasmid (construct ΔSst) or in the construct ΔDo (construct ΔDoΔSst, ΔDoΔKpn1, and ΔDoΔKpn2). More than 95% of the RNA molecules produced from constructs ΔSst, ΔDoΔSst, and ΔDoΔKpn2 were processed around the CaMV site (Fig. 4). However, the 3′ ends of those RNAs showed more heterogeneity than RNAs produced from the standard plasmid (see Fig. 5). RNAs produced from construct ΔDoΔKpn1 also showed 3′-end

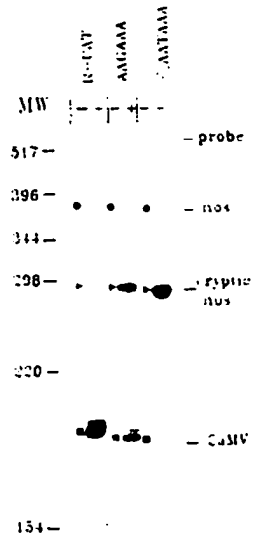


Figure 3. Protection assay of RNAs produced from the AATAAA mutants with probe P1. Lane MW indicates the position of relative molecular weight markers (^{32}P-labeled single-strand DNA fragments). For each construct the sense–antisense hybrids are shown before (−) or after (+) digestion with RNase A and T1. Antisense probes used were the exact mutant derivatives of probe P1 (Fig. 2). Expected sizes of transcripts processed at the CaMV (■), the cryptic (▲), and the normal (●) *nos* sites are shown for each lane.

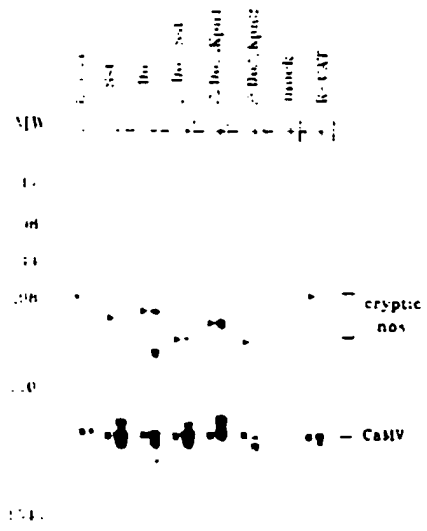


Figure 4. Protection assay with RNAs produced from mutants in the downstream region with probe P1. (MWI Relative molecular weight markers (32P-labeled single-stranded DNA fragments). Hybrids before (−) and after (+) digestion are shown. The antisense probes used are mutant derivatives of probe P1. The expected sizes for transcripts processed at the CaMV (■) or at the cryptic *nos* (▲) sites are indicated for each lane.

heterogeneity at the CaMV site and some readthrough to the cryptic *nos* site (12% of the transcripts). Taken together, these results suggest that in contrast to most reported higher eukaryotic systems, the primary sequence downstream of the CaMV processing site does not contain elements necessary for 3′-end formation, although the primary or the secondary structure in the immediate surrounding of the poly(A) site might influence the precise positioning of cleavage.

Sequences upstream of AATAAA are important for CaMV RNA 3′-end formation

To localize sequences important for processing, we progressively deleted the sequences upstream from the AATAAA by BAL-31 digestion from the 5′ end of the CaMV poly(A) signal fragment toward the AATAAA sequence (Fig. 2A). The mutants obtained were named according to the number of nucleotides remaining upstream of the processing site (as shown in Fig. 1). The sequence at the junction between the CAT gene and the truncated CaMV signal varied slightly from one clone to another, as indicated in Materials and methods. RNAs were analyzed using probe P1 (see Fig. 2). The length of the protected fragments for transcripts processed at the CaMV, the cryptic, and the normal *nos* poly(A) addition site decreases with increasing deletion (Fig. 6A). Transcripts processed at the CaMV site in mutants with the largest deletions (construct Δ75–Δ32) would be very small and therefore difficult to quantitate with our procedure. To increase the size of the protected antisense fragments, we therefore used homologous probes (probe series P1-1) for each mutant complementary to the end of the CAT gene, the mutated CaMV signal, and the rescuing *nos* signal (Figs. 2A and 6B).

On progressive deletion of upstream sequences, the efficiency of cleavage at the CaMV poly(A) addition site is reduced. The sequences upstream of nucleotide 67, which include two TG-rich motifs (Fig. 1), are likely to play only an auxiliary role in mRNA 3′-end formation since 60% of the transcripts produced from construct Δ67 were processed at the CaMV site (Fig. 6). In particular, deletion of the second TG-rich sequence 70 nucleotides upstream of the processing site did not affect processing efficiency (cf. constructs Δ75 and Δ67, Fig. 6B). Interestingly, the transcripts that bypassed the CaMV signal are mainly processed at the cryptic rather than at the normal *nos* site. Deletion of an additional 24 nucleotides reduced processing at the CaMV site to 46% (construct Δ44). The transcripts that read through the CaMV signal were processed at either the cryptic or the normal *nos* site. Finally, deletion of the subsequent 12 nucleotides allowed only very limited cleavage at the CaMV site (8%, construct Δ32). Most of the remaining transcripts in this case were processed at the normal *nos* site. These results suggest the involvement of at least one and perhaps several upstream elements in the recognition of AATAAA, either CaMV or *nos*. The most important element spans the region between 67 and 32 nucleotides upstream of the CaMV processing site (Fig. 1).

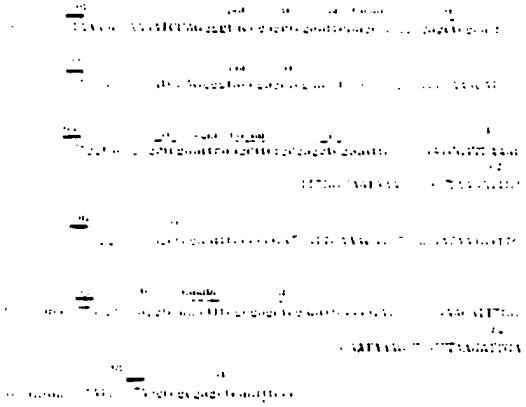
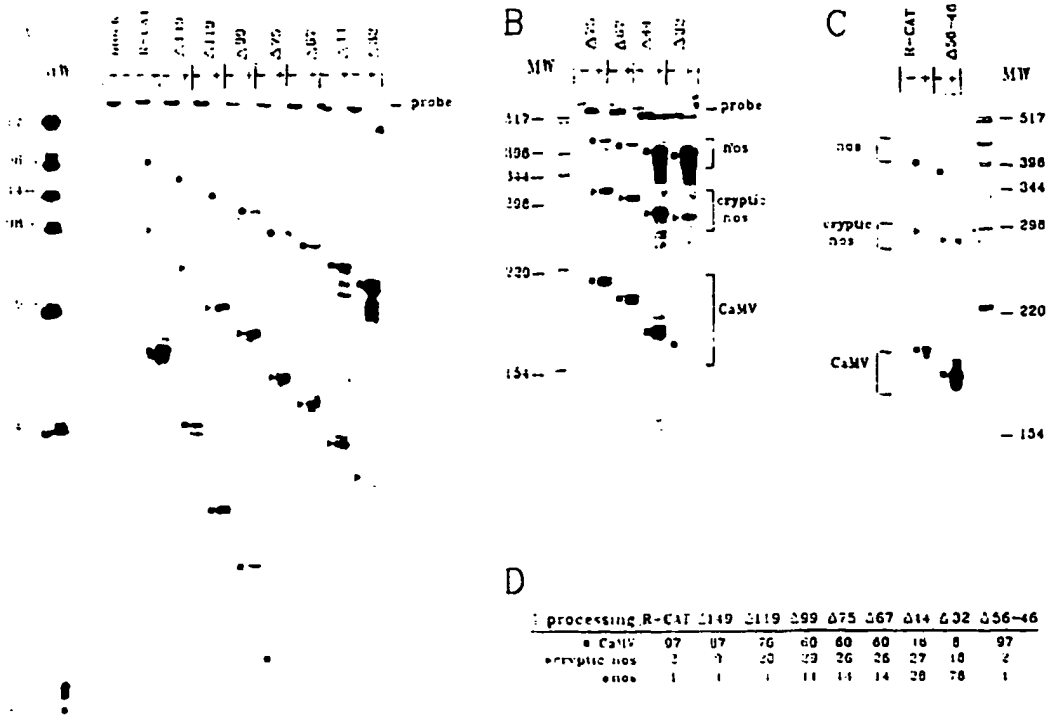


Figure 5. Primary structure of mutants in the downstream region. Uppercase letters represent the original CaMV or *nos* sequences; lowercase letters represent polylinker sequences. The wild-type CaMV processing site is shown by the solid box above each sequence. Location of additional processing sites in the downstream mutants was evaluated according to the size of the protected fragment and is indicated by the open boxes. Numbers above those boxes represent the percentage of transcripts cleaved at the corresponding site. The cryptic AATAAA in the *nos* signal is enclosed in a dotted box.



D

processing	R-CAT	Δ149	Δ119	Δ99	Δ75	Δ67	Δ14	Δ32	Δ56–46
• CaMV	97	97	76	60	60	60	18	8	97
▲cryptic nos	2	3	20	29	26	26	27	18	2
●nos	1	1	1	11	14	14	28	78	1

Figure 6. Protection assays with RNAs produced from mutants in the upstream CaMV region. (*A*) Analysis of the BAL-31 deletion mutants with probe P1. (MW) Molecular weight markers. Hybrids before (−) and after (+) RNase digestion are shown. The expected sizes of transcripts processed at the CaMV (■), the cryptic (▲), or the normal (●) *nos* sites are shown for each lane. (*B*) Analysis of the four largest BAL-31 deletion mutants with probe P1-1. For each mutant, the P1-1 probe used is a derivative of the wild-type P1-1 with the corresponding junction between the CAT gene and the deleted CaMV signal. As in *A*, expected sizes of transcripts processed at each site are indicated. (*C*) Analysis of the precise deletion mutant Δ56–46 using probe P1. For analysis of the mutant, a derivative of probe P1 containing the same mutation in the CaMV signal was used. (*D*) Percentage of transcripts processed at each site. Numbers presented are a compilation of several experiments obtained with either probe P1 or probe P1-1.

A precise deletion of the sequence CCCTTAGTATG situated 56–46 nucleotides upstream of the processing site (construct Δ56–46) did not affect processing at the CaMV site (Fig. 6C), suggesting that the upstream element mentioned above might be restricted to 45 nucleotides upstream of the processing site. These sequences include perfect repeats of the motif TATTTGTA.

To determine which of the CaMV sequences induced recognition of the normally silent *nos* AATAAA in several of our deletion mutants, we inserted two types of oligonucleotides containing part of the CaMV upstream sequence, in either orientation, 35 nucleotides upstream of the cryptic and 150 nucleotides upstream of the normal *nos* poly(A) addition site in plasmid *nos* (Fig. 2B). Oligonucleotide 53–32, which includes sequences 53–32 nucleotides upstream of the CaMV processing site (TTAGTATGTATTTGTATTTGTA), and oligonucleotide 76–70, which includes sequences 76–70 nucleotides upstream of the CaMV processing site (TGTGTTG), were chosen (Fig. 1; see also Materials and methods). The RNAs produced from these constructs were analyzed using homologous probes Pnos, which covered for each mutant the *nos* fragment, the introduced oligonucleotide, and the end of the CAT gene (Fig. 2B). In the parent plasmid, processing occurred only at the normal *nos* site (Fig. 7, construct *nos*). Introduction of oligonucleotide 53–32 induced the recognition of the cryptic AATAAA, since 30% of the transcripts were cleaved at the cryptic site (Fig. 7, construct *nos*+(53–32)). This element had no effect when placed in the reverse orientation (Fig. 7, construct *nos*+(32–53)). The sequences contained in oligonucleotide 76–70 did not induce the AATAAA recognition in either orientation. The active oligonucleotide (53–32) includes tandem repeats of the TATTTGTA motif mentioned above.

Discussion

We have defined several elements involved in the 3′-end formation of CaMV RNA. Interestingly, the requirements for this plant poly(A) signal seem to differ from what is known for vertebrate systems. To our knowl-

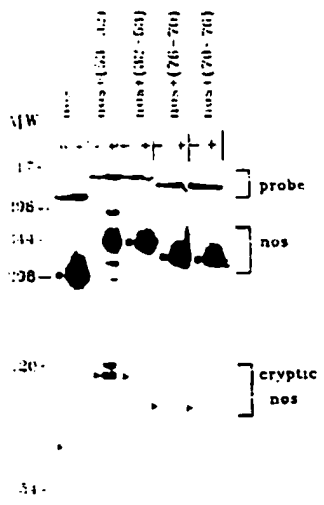


Figure 7. Protection assay, with probe Pnos, of RNAs produced from the CaMV oligonucleotide insertions upstream of the *nos* poly(A) signal. The mapping strategy is shown in Fig. 2B. (MW) Molecular weight markers; (−) and (+) hybrids before or after RNase digestion. Expected size of protected fragments corresponding to transcripts processed at the normal (●) and the cryptic (▲) *nos* sites are shown for each lane.

edge, the results presented here and earlier (Sanfaçon and Hohn 1990) provide the first direct evidence that an AATAAA sequence is important for mRNA 3′-end formation in plant cells. We also show that, in contrast to vertebrate systems, a point mutation in this sequence is partly recognized. It is noteworthy that in vertebrate in vivo (Montell et al. 1983) and in vitro (Wilusz et al. 1989) systems, a similar mutation would reduce cleavage by >95%. Apparently, the stringency for the AATAAA consensus sequence is much lower than in vertebrate systems. This is in agreement with the previous observation that a perfect consensus AATAAA sequence upstream of the processing site is present in only 40% of known plant genes (Joshi 1987). The normal *nos* processing site, for instance, is not preceded by a perfect consensus AATAAA (Fig. 1; Bevan et al. 1982; Depicker et al. 1982). It is very likely that sequences other than AATAAA play an important role and can improve the recognition of less conserved AATAAA in plant genes.

The sequences downstream of the CaMV poly(A) site do not influence processing efficiency drastically, although they affect the precision of cleavage. Similar precision effects were observed with vertebrate systems using minor mutations in the downstream elements (Woychick et al. 1984; Mason et al. 1986). The heterogeneity of processing observed with some of our mutants can be due to either the primary or the secondary structure of sequences surrounding the cleavage site. The fragment containing the wild-type CaMV and *nos* poly(A) signals can be folded into a potential structure consisting of several stems and loops. The processing site is situated just upstream of an AC-rich region that is relatively unstructured in our model.

In contrast to CaMV, TG-rich regions are found downstream of the cleavage site of vertebrate genes and are essential for formation of the polyadenylation complexes and for efficiency of the cleavage and polyadenylation reactions (for review, see Humphrey and Proudfoot 1988). In vitro, the recognition of a normally silent AATAAA was induced on downstream insertion of an oligonucleotide containing the TG-rich sequence (Ryner et al. 1989). In plant systems, deletion of TG-rich sequences downstream of the poly(A) site of the octopine synthetase (Ingelbrecht et al. 1989) and the potato wound-inducible proteinase inhibitor II (An et al. 1989) genes reduced steady-state RNA level or reporter gene activity. The latter results would, however, not allow one to distinguish between effects at the level of cleavage and polyadenylation efficiency, of differences in mRNA stability, or of transport efficiency. Work with additional plant poly(A) signals will give information as to whether the absence of a positively acting downstream element (such as a T-rich or TG-rich element) is a feature of more plant poly(A) signals or a peculiarity of the CaMV.

We could show that deletion of sequences upstream of CaMV AATAAA inhibits processing at the CaMV site. Furthermore, the recognition of a normally silent AATAAA was induced on insertion upstream of oligonucleotide 53–32, which includes the sequence TTAGTATGTATTTGTATTTGTA (Fig. 7). The induced recognition of this cryptic AATAAA was, however, partial (33% of the transcripts; Fig. 7). One possible explanation of this partial recognition is that we introduced only one of several CaMV upstream elements, which perhaps normally act in an additive fashion. This would be consistent with the results obtained with mutant Δ67, which contained the whole CaMV sequence of oligonucleotide 53–32 but showed only partial cleavage at the CaMV site (60% of the transcripts). Alternatively, the positive effect of the upstream element may be dependent on its position relative to AATAAA. In wild-type CaMV, this element is situated 14 nucleotides upstream of AATAAA, while in the artificial situation it is 30 nucleotides upstream of the cryptic *nos* AATAAA. Such a position-dependent action of downstream elements has been described in vertebrate cells (Gil and Proudfoot 1987; Heath et al. 1990).

Because the oligonucleotide 53–32 alone has a positive effect on insertion into a heterologous signal, it is likely that its primary structure rather than the general secondary structure of the CaMV upstream region is important. The CaMV element contained in this oligonucleotide has homologies with proposed consensus sequences in vertebrate poly(A) signals (TG-rich downstream element) and with a proposed consensus sequence TGTGTTT found downstream of some plant-processing sites (Joshi 1987). The other tested sequence,

TGTGTTG, which is a single, short TG-rich element, did not have any influence on either CaMV or cryptic *nos* AATAAA recognition. It is possible that the 53–32 upstream sequence is active because of the presence of several repeats of a short TG-rich sequence (such as TTTGTA or TGTATTT; see Fig. 1). Similarly, it was shown that the downstream element of the rabbit β-globin gene consisted of several smaller elements (Gil and Proudfoot 1987). It is also possible that the important feature of the 53–32 sequence is not simply TG richness. It is interesting that this sequence also shows homologies to a yeast terminator consensus sequence (TAG TAGT . . . TTT; Zaret and Sherman 1982). However, this homology is perhaps not relevant since a precise deletion of the first part of this sequence in the CaMV signal allowed processing at wild-type efficiency (Fig. 6, construct Δ56–46). Furthermore, more recent experiments suggest that this sequence might not be a key element in yeast mRNA 3′-end processing (Osborne and Guarente 1989). Point mutation analysis of CaMV upstream element(s) is under way, which will allow us to determine the important features of this region.

The involvement of sequences upstream of AATAAA has been described in only a few systems. In plants, the analysis of the pea ribulose-1′5-biphosphate carboxylase small-subunit gene suggested the involvement of several upstream and downstream elements (Hunt and MacDonald 1989). In vertebrate systems, it has been shown that an AATAAA and a downstream element are the only necessary requirements for efficient processing and that the polyadenylation site is usually at an A residue (Kessler et al. 1986; Levitt et al. 1989). However, upstream elements in addition to the downstream element have been described for the SV40 late poly(A) signal (Carswell and Alwine 1989), for the adenovirus late transcription unit (DeZasso and Imperiale 1989), and for the hepatitis B virus (Russnak and Ganem 1990). In most cases, the upstream element induces the recognition of weak poly(A) signals, thereby allowing a tight regulation of mRNA 3′-end formation. In the case of the hepatitis B virus, a pararetrovirus such as CaMV, part of this element is upstream of the transcription initiation site and allows the production of terminally redundant RNA. In the case of CaMV, the conditional recognition of the poly(A) signal is dependent on promoter proximity (Sanfaçon and Hohn 1990) and not on the upstream element, or elements, which is located downstream of the transcription initiation site. A possible regulatory role for an upstream element in CaMV remains unclear. Since the element comprised in oligonucleotide 53–32 stimulates AATAAA recognition in only one orientation, it is tempting to speculate that it acts at the RNA level, although we cannot rule out the possibility that it acts at the DNA level. It is possible that this upstream element is involved in the formation of a polyadenylation complex and compensates for the absence of a downstream element. Whether the involvement of upstream elements is a particular feature of a few poly(A) signals (e.g., those originating from viruses or from plants) or a more general feature remains to be seen.

Materials and methods

Plasmid constructions

Construction of the plasmid R-CAT was described previously (Sanfaçon and Hohn 1990). Plasmids ΔAATAAA, AAGAAA, ΔDo, and Δ56–46 were obtained by oligonucleotide-directed mutagenesis of plasmid R-CAT. Plasmids ΔSst and ΔDoΔSst were obtained by cutting plasmids R-CAT and ΔDo, respectively, with *Sst*I and religating the large isolated fragment. Plasmids ΔDoΔKpn1 and ΔDoΔKpn2 were obtained serendipitously by cutting plasmid ΔDo with *Kpn*I and *Hin*dIII and religating; the relevant sequences are shown in Figure 5. Plasmids Δ149–Δ32 were obtained by digestion of plasmid R-CAT with *Pst*I between the CAT gene and the CaMV poly(A) signal fragment. The linearized fragment was digested further with BAL-31 and ligated to either an *Xho*I or a *Pst*I linker. The mixture was digested with either *Xho*I–*Hin*dIII or *Pst*I–*Hin*dIII, and the small fragment containing the deleted CaMV poly(A) signal was isolated and religated with the large *Pst*I–*Hin*dIII fragment of plasmid R-CAT. The resulting sequences between position −196 (Fig. 1) and the deletion point are as follows: plasmids Δ149–Δ75, GAAGCTTCTCGAGC; plasmid Δ67, GAAGCTC; plasmid Δ44, GAAGCTCC; and plasmid Δ32, GAAGCTCCTCGAGG. Plasmid *nos* was constructed by ligating the small *Pst*I–*Hin*dIII fragment of pNOSCAT (Fromm et al. 1985) containing the *nos* poly(A) signal into the large *Pst*I–*Hin*dIII fragment of pDW2 (Pietrzak et al. 1986). Plasmid *nos*+(53–32) was constructed by inserting oligonucleotide 53–32: 5′-CTGCAGACTCGAGTTAGTATGTATTTGTATTTGTAGTCGACACTGCAG-3′ into the *Pst*I site of plasmid *nos*. This oligonucleotide contains the CaMV sequence 53–32 nucleotides upstream of the processing site surrounded by polylinker sequences: a *Pst*I and a *Xho*I site on one extremity and a *Pst*I and a *Sal*I site on the other extremity. Plasmid *nos*+(32–53) was obtained by cloning the same oligonucleotide in the reverse orientation. Plasmid *nos*+(76–70) and *nos*+(70–76) were built in a similar manner using oligonucleotide 70–76: 5′-CTGCAGACTCGAGTCTGTTGGTCGACACTGCAG-3′, which includes the relevant CaMV sequence surrounded by the same polylinker sequences. Probe P1 has been described (Sanfaçon and Hohn 1990). For each mutant a derivative of this probe was constructed by inserting the small *Pst*I–*Eco*RI fragment containing the mutated CaMV poly(A) signal and the wild-type *nos* poly(A) signal into probe P1. Probes P1-1 were obtained for each mutant by transferring the corresponding *Bam*HI–*Hin*dIII fragment containing the CAT gene and the CaMV poly(A) signal into the *Pvu*II–*Hin*dIII fragment of probe P1. Similarly, probe Pnos was obtained by transferring the *Bam*HI–*Hin*dIII fragment of plasmid *nos* containing the CAT gene and the *nos* poly(A) signal into the *Pvu*II–*Hin*dIII fragment of probe P1. Derivatives of probe Pnos were obtained for the oligonucleotide insertion mutants.

Protoplast transfection and RNA analysis

N. plumbaginifolia protoplasts were transfected, and total RNA was purified after 6 hr and mapped by RNase protection assay as described previously (Vankan et al. 1988; Goodall and Filipowicz 1989). For each of the mutants tested, a specific antisense probe containing the same mutation was used. Probes P1 were digested with *Pst*I prior to in vitro transcription, and probes P1-1 and Pnos were digested with *Sca*I (site located at the end of the CAT gene). For the experiments using probe Pnos and derivatives (Fig. 7) the sense–antisense hybrids were digested by RNase T1 only. This procedure avoided the appearance of secondary bands due to breathing of the hybrids in a region very